

NONPROVISIONAL

PATENT APPLICATION

METHODS FOR IMPROVED ARRAY PREPARATION

Inventors:

Glenn McGall, a citizen of Canada , residing at, 750 No. Shoreline Blvd. #30,
Mountain View, California, 94043

Lisa Kajisa, a citizen of the United States, residing at 4644 Doyle Court, San
Jose, California 95129

Mark Trulson, a citizen of the United States, residing at 1240 Martin Ave.,
San Jose California, 95126

Assignee:

Affymetrix, Inc.
3380 Central Expressway
Santa Clara, CA 95051

Entity: Large

METHODS FOR IMPROVED ARRAY PREPARATION

CROSS-REFERENCES TO RELATED APPLICATIONS

5 This application claims priority from USSN 60/174,980 filed January 6, 2000, the disclosure of which is incorporated herein by reference.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

10 Not applicable

BACKGROUND OF THE INVENTION

15 The present invention relates to improved methods for preparing support-bound nucleic acid arrays. More particularly, the invention relates to methods of preparing and packaging the arrays wherein agents that can affect the variability and performance of the arrays are excluded from the environment.

20 Substrate-bound nucleic acid arrays, such as the Affymetrix DNA Chip, enable one to test hybridization of a target nucleic acid molecule to many thousands of differently sequenced nucleic acid probes at feature densities greater than about five hundred per 1 cm². Because hybridization between two nucleic acids is a function of their sequences, analysis of the pattern of hybridization provides information about the sequence of the target molecule. The technology is useful for *de novo* sequencing and re-sequencing of nucleic acid molecules and also has important diagnostic uses in
25 discriminating genetic variants that may differ in sequence by one or a few nucleotides. For example, substrate-bound nucleic acid arrays are useful for identifying genetic variants of infectious diseases, such as HIV, or genetic diseases, such as cystic fibrosis.

30 In one version of the substrate-bound nucleic acid array, the target nucleic acid is labeled with a detectable marker, such as a fluorescent molecule. Hybridization between a target and a probe is determined by detecting the fluorescent signal at the various locations on the substrate. The amount of signal is a function of the thermal stability of the hybrids. The thermal stability is, in turn, a function of the sequences of the target-probe pair: AT-rich regions of DNA melt at lower temperatures than GC-rich

regions of DNA. This differential in thermal stabilities is the primary determinant of the breadth of DNA melting transitions, even for nucleic acids.

Depending upon the length of the nucleic acid probes, the number of different probes on a substrate, the length of the target nucleic acid, and the degree of

5 hybridization between sequences containing mismatches, among other things, a hybridization assay carried out on a substrate-bound nucleic acid array can generate thousands of data points of different signal strengths that reflect the sequences of the probes to which the target nucleic acid hybridized. This information can require a computer for efficient analysis. The fact of differential fluorescent signal due to
10 differences in thermal stability of hybrids complicates the analysis of hybridization results, especially from combinatorial nucleic acid arrays for *de novo* sequencing and custom nucleic acid arrays for specific re-sequencing applications. Modifications in custom array designs have contributed to simplifying this problem.

Further complications can arise and lead to variability in diagnostic or
15 sequencing results. For example, degradation of nucleic acid probes, either during the synthesis steps or on standing can lead to variability in assay results. Accordingly, there exists a need for additional methods of nucleic acid array preparation, and the arrays themselves, to provide more robust tools for the skilled researcher. The present invention provides such methods and arrays.

20

SUMMARY OF THE INVENTION

In one aspect, the present invention provides methods for preparing nucleic acid arrays on a support. In these methods a plurality of nucleic acids are
25 synthesized on the support and the synthesis steps are carried out in a non-oxidizing atmosphere.

In one group of embodiments, each nucleic acid occupies a separate predefined region of the support, the synthesizing comprising:

- (a) activating a region of the support;
- 30 (b) attaching a nucleotide to a first region, the nucleotide having a masked reactive site linked to a protecting group;
- (c) repeating steps (a) and (b) on other regions of the support whereby each of the other regions has bound thereto another nucleotide comprising a masked

reactive site linked to a protecting group, wherein the other nucleotide may be the same or different from that used in step (b);

(d) removing the protecting group from one of the nucleotides bound to one of the regions of the support to provide a region bearing a nucleotide having an

5 unmasked reactive site;

(e) binding an additional nucleotide to the nucleotide with an unmasked reactive site;

(f) repeating steps (d) and (e) on regions of the support until a desired plurality of nucleic acids is synthesized, each nucleic acid occupying separate predefined regions of the support;

10 wherein each of steps (a) through (f) are carried out in a non-oxidizing atmosphere.

In another group of embodiments, the synthesizing comprises the sequential steps of:

15 a) generating a pattern of light and dark areas by selectively irradiating at least a first area of a surface of a substrate, the substrate comprising immobilized nucleotides on the surface, and the nucleotides capped with a photoremovable protective group, without irradiating at least a second area of the surface, to remove the protective group from the nucleotides in the first area;

20 b) simultaneously contacting the first area and the second area of the surface with a first nucleotide to couple the first nucleotide to the immobilized nucleotides in the first area, and not in the second area, the first nucleotide capped with the photoremovable protective group;

25 c) generating another pattern of light and dark areas by selectively irradiating with light at least a part of the first area of the surface and at least a part of the second area to remove the protective group in at least a part of the first area and at least a part of the second area;

d) simultaneously contacting the first area and the second area of the surface with a second nucleotide to couple the second nucleotide to the immobilized nucleotides in at least a part of the first area and at least a part of the second area;

30 e) performing additional irradiating and nucleotide contacting and coupling steps so that a matrix array of at least 100 nucleic acids having different sequences is formed on the support;

with the proviso that steps (a) through (e) are performed in a non-oxidizing atmosphere, and the atmosphere has an ozone concentration of from about 0 to about 5 ppb. Preferably, the atmosphere is filtered air or an inert gas, more preferably argon.

~~In another aspect, the present invention provides a method of preparing~~

5 and packaging a nucleic acid array, comprising:

(a) preparing the array by any of the methods noted above; and

(b) packaging the array prepared in (a) in an enclosure having a non-oxidizing atmosphere. In preferred embodiments, the nucleic acid array prepared in step (a) is exposed to unfiltered air for a period of two hours or less, before packaging. More
10 preferably, the entire process of preparing and packaging the array is conducted in a non-oxidizing atmosphere.

In yet another aspect, the present invention provides a method of preparing a nucleic acid array, comprising attaching each of a plurality of nucleic acids to a solid support at preselected locations, wherein the attaching is carried out in a non-oxidizing
15 atmosphere.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the effect of unfiltered air on the degradation of several
20 nucleic acid probes. In Figure 1A, the probes are exposed to unfiltered air then cleaved from the support and analyzed by HPLC. In Figure 1B, the probes are exposed to an ozone atmosphere (1000 ppb) for the indicated time, cleaved from the support and analyzed by HPLC.

Figure 2 illustrates the effect of ozone over time on support bound nucleic
25 acids.

Figure 3 illustrates the effect of total ozone concentration on support-bound nucleic acid degradation.

Figure 4 illustrates the effect of various environments on the degradation of support-bound nucleic acid probes. The probes were exposed to purified air, nitrogen,
30 oxygen and ambient air.

Figure 5 illustrates the yield increase for preparation of a 16-mer nucleic acid probe that can be achieved using photolithography with backside illumination.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

The following definitions are set forth to illustrate and define the meaning

5 and scope of the various terms used to describe the invention herein.

"Nucleic acid library" or "array" is an intentionally created collection of nucleic acids which can be prepared either synthetically or biosynthetically and screened for biological activity in a variety of different formats (*e.g.*, libraries of soluble molecules; and libraries of oligos tethered to resin beads, silica chips, or other solid supports).

10 Additionally, the term "array" is meant to include those libraries of nucleic acids which can be prepared by spotting nucleic acids of essentially any length (*e.g.*, from 1 to about 1000 nucleotide monomers in length) onto a substrate. The term "nucleic acid" as used herein refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides, that comprise purine and pyrimidine bases, or other natural,
15 chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The backbone of the polynucleotide can comprise sugars and phosphate groups, as may typically be found in RNA or DNA, or modified or substituted sugar or phosphate groups. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. The sequence of nucleotides may be interrupted by non-nucleotide
20 components. Thus the terms nucleoside, nucleotide, deoxynucleoside and deoxynucleotide generally include analogs such as those described herein. These analogs are those molecules having some structural features in common with a naturally occurring nucleoside or nucleotide such that when incorporated into a nucleic acid or oligonucleoside sequence, they allow hybridization with a naturally occurring nucleic
25 acid sequence in solution. Typically, these analogs are derived from naturally occurring nucleosides and nucleotides by replacing and/or modifying the base, the ribose or the phosphodiester moiety. The changes can be tailor made to stabilize or destabilize hybrid formation or enhance the specificity of hybridization with a complementary nucleic acid sequence as desired.

30 "Solid support", "support", and "substrate" are used interchangeably and refer to a material or group of materials having a rigid or semi-rigid surface or surfaces. In many embodiments, at least one surface of the solid support will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different compounds with, for example, wells, raised regions, pins, etched

trenches, or the like. According to other embodiments, the solid support(s) will take the form of beads, resins, gels, microspheres, or other geometric configurations.

"Predefined region" or "preselected region" refers to a localized area on a solid support which is, was, or is intended to be used for formation of a selected molecule

5 and is otherwise referred to herein in the alternative as a "selected" region, a "known" region, or a "known" location. The predefined or known region may have any convenient shape, *e.g.*, circular, rectangular, elliptical, wedge-shaped, etc. For the sake of brevity herein, "predefined regions" are sometimes referred to simply as "regions." In some embodiments, a predefined or known region and, therefore, the area upon which each
10 distinct compound is synthesized is smaller than about 1 cm² or less than 1 mm². Within these regions, the molecule synthesized therein is preferably synthesized in a substantially pure form. In additional embodiments, a predefined region can be achieved by physically separating the regions (*i.e.*, beads, resins, gels, etc.) into wells, trays, etc. Accordingly, materials (*e.g.*, nucleic acids) can be synthesized or attached to any particular region by
15 any known methods or means.

As used herein, the term "non-oxidizing atmosphere" refers to an atmosphere or environment that produces less than 30% degradation of a nucleic acid over a period of about 1 hour at ambient temperature (about 18 to about 25°C). More preferably, a non-oxidizing atmosphere produces less than 20% degradation, still more
20 preferably less than 10% degradation. Most preferably, a non-oxidizing atmosphere produces less than 5% degradation of a nucleic acid under the conditions noted above. Evaluation of a particular atmosphere for nucleic acid or probe degradation can be carried out using a support-bound nucleic acid heteropolymer having from 8-16 residues and attached to the support by a cleavable linking group. After the indicated exposure to the
25 atmosphere selected, the nucleic acid can be cleaved from the support and subjected to HPLC analysis relative to a standard (nucleic acid exposed to an inert atmosphere, for example, argon). Alternatively, a control array and a test array can be exposed to an atmosphere of interest and then evaluated for DNA binding to determine an amount of degradation.

30

General

Nucleic acid arrays having single-stranded nucleic acid probes have become powerful research tools for identifying and sequencing new genes. Other arrays

of unimolecular double-stranded DNA have been developed which are useful in a variety of screening assays and diagnostic applications (see, for example, U.S. Patent No. 5,556,752). Still other arrays have been described in which a ligand or probe (a peptide, for example), is held in a conformationally restricted position by two complementary

5 nucleic acid, at least one of which is attached to a support. Common to each of these types of arrays is the presence of a support-bound nucleic acid and the exquisite sensitivity exhibited by the arrays. Unfortunately, the sensitivity of these arrays can be compromised if the nucleic acids are degraded.

10 In order to provide the researcher with arrays of uncompromising quality and reproducible performance, arrays should be prepared using high yield reactions and excluding any component which could negatively impact synthesis yield or the performance of the array.

The present invention derives from the discovery that improved yields and reduced product variability can be obtained if nucleic acid arrays are prepared in a non-oxidizing atmosphere, typically an atmosphere which is essentially free of ozone.
15 Exposure of solid supports to unfiltered air during the synthesis process of preparing nucleic acid arrays can cause a reduction in yield. By conducting the synthesis steps in a controlled atmosphere such as filtered air or an inert gas (*e.g.*, argon), a 5-10 fold increase in nucleic acid probe yields can be obtained. Additionally, once the nucleic acid arrays
20 are prepared, degradation can occur during post-synthesis processing steps if the arrays are exposed to unfiltered air. Storage of the arrays under nitrogen protects the arrays from such degradation.

Accordingly, the present invention provides methods wherein nucleic acid arrays are prepared in a non-oxidizing atmosphere. Additionally, the present invention
25 provides methods wherein the arrays are packaged in a non-oxidizing atmosphere, and wherein handling of arrays prior to packaging is carried out in a manner which limits the array's exposure to unfiltered air.

Embodiments of the Invention

30

In view of the above discoveries, the present invention provides an improved method of preparing a nucleic acid array on a support. In a general sense, the method comprises synthesizing a plurality of nucleic acids on a support wherein the synthesis steps are carried out in a non-oxidizing atmosphere.

Synthesis of Nucleic acid Arrays

In the present invention, nucleic acid arrays can be prepared using a variety of synthesis techniques directed to high-density arrays of nucleic acids on solid supports. In brief, the methods can include light-directed methods, flow channel or spotting methods, pin-based methods, bead-based methods or combinations thereof. For light-directed methods, see, for example, U.S. Patent No. 5,143,854, 5,424,186 and 5,510,270. For techniques using mechanical methods, see PCT No. 92/10183, U.S. Patent No. 5,384,261 and PCT/US99/00730. For a description of bead based techniques, see PCT US/93/04145, and for pin-based methods, see U.S. Patent No. 5,288,514. A brief description of these methods is provided below. The methods of the present invention are equally amenable to the preparation of unimolecular double-stranded DNA arrays (see U.S. Patent No. 5,556,752). In addition, the nucleic acid arrays prepared in the present methods will also include those arrays in which individual nucleic acids are interrupted by non-nucleotide portions (see, for example U.S. Patent No. 5,556,752 in which probes such as polypeptides are held in a conformationally restricted manner by complementary nucleic acid fragments).

Various additional techniques for large scale polymer synthesis are known. Some examples include the U.S. Patents Nos.: 5,143,854, 5,242,979, 5,252,743, 5,324,663, 5,384,261, 5,405,783, 5,412,087, 5,424,186, 5,445,934, 5,451,683, 5,482,867, 5,489,678, 5,491,074, 5,510,270, 5,527,681, 5,550,215, 5,571,639, 5,593,839, 5,599,695, 5,624,711, 5,631,734, 5,677,195, 5,744,101, 5,744,305, 5,753,788, 5,770,456, 5,831,070, and 5,856,011, all of which are incorporated by reference herein.

Libraries on a Single Substrate

Light-Directed Methods

For those embodiments using a single solid support, the nucleic acids of the present invention can be formed using techniques known to those skilled in the art of polymer synthesis on solid supports. Preferred methods include, for example, "light directed" methods which are one technique in a family of methods known as VLSIPSTM methods. The light directed methods discussed in U.S. Patent No. 5,143,854 involve activating predefined regions of a substrate or solid support and then contacting the

substrate with a preselected monomer solution. The predefined regions can be activated with a light source, typically shown through a mask (much in the manner of photolithography techniques used in integrated circuit fabrication). Other regions of the substrate remain inactive because they are blocked by the mask from illumination and remain chemically protected. Thus, a light pattern defines which regions of the substrate react with a given monomer. By repeatedly activating different sets of predefined regions and contacting different monomer solutions with the substrate, a diverse array of nucleic acids is produced on the substrate. Of course, other steps such as washing unreacted monomer solution from the substrate can be used as necessary.

The VLSIPSTM methods are preferred for the methods described herein. Additionally, the surface of a solid support, optionally modified with spacers having photolabile protecting groups such as NVOC and MeNPOC, is illuminated through a photolithographic mask, yielding reactive groups (typically hydroxyl groups) in the illuminated regions. A 3'-O-phosphoramidite activated deoxynucleoside (protected at the 5'-hydroxyl with a photolabile protecting group) is then presented to the surface and chemical coupling occurs at sites that were exposed to light. Following capping, and oxidation, the substrate is rinsed and the surface illuminated through a second mask, to expose additional hydroxyl groups for coupling. A second 5'-protected, 3'-O-phosphoramidite activated deoxynucleoside is presented to the surface. The selective photodeprotection and coupling cycles are repeated until the desired set of nucleic acids is produced. Alternatively, an oligomer of from, for example, 4 to 30 nucleotides can be added to each of the preselected regions rather than synthesize each member in a monomer by monomer approach. Methods for light-directed synthesis of DNA arrays on glass substrates are also described in McGall *et al.*, *J. Am. Chem. Soc.*, **119**:5081-5090 (1997).

For the above light-directed methods wherein photolabile protecting groups and photolithography are used to create spatially addressable parallel chemical synthesis of a nucleic acid array (see also U.S. Patent No. 5,527,681), computer tools may be used to assist in forming the arrays. For example, a computer system may be used to select nucleic acid or other polymer probes on the substrate, and design the layout of the array as described in, for example, U.S. Patent No. 5,571,639.

Flow Channel or Spotting Methods

Additional methods applicable to library synthesis on a single substrate are described in U.S. Pat. No. 5,384,261 and in PCT/US99/00730. In the methods disclosed in this patent and PCT publication, reagents are delivered to the substrate by either (1)

5 flowing within a channel defined on predefined regions or (2) "spotting" on predefined regions. However, other approaches, as well as combinations of spotting and flowing, may be employed. In each instance, certain activated regions of the substrate are mechanically separated from other regions when the monomer solutions are delivered to the various reaction sites.

10 A typical "flow channel" method applied to the compounds and libraries of the present invention can generally be described as follows. Diverse nucleic acid sequences are synthesized at selected regions of a substrate or solid support by forming flow channels on a surface of the substrate through which appropriate reagents flow or in which appropriate reagents are placed. For example, assume a monomer "A" is to be
15 bound to the substrate in a first group of selected regions. If necessary, all or part of the surface of the substrate in all or a part of the selected regions is activated for binding by, for example, flowing appropriate reagents through all or some of the channels, or by washing the entire substrate with appropriate reagents. After placement of a channel block on the surface of the substrate, a reagent having the monomer A flows through or is
20 placed in all or some of the channel(s). The channels provide fluid contact to the first selected regions, thereby binding the monomer A on the substrate directly or indirectly (via a spacer) in the first selected regions.

Thereafter, a monomer B is coupled to second selected regions, some of which may be included among the first selected regions. The second selected regions will
25 be in fluid contact with a second flow channel(s) through translation, rotation, or replacement of the channel block on the surface of the substrate; through opening or closing a selected valve; or through deposition of a layer of chemical or photoresist. If necessary, a step is performed for activating at least the second regions. Thereafter, the monomer B is flowed through or placed in the second flow channel(s), binding monomer
30 B at the second selected locations. In this particular example, the resulting sequences bound to the substrate at this stage of processing will be, for example, A, B, and AB. The process is repeated to form a vast array of sequences of desired length at known locations on the substrate.

After the substrate is activated, monomer A can be flowed through some of the channels, monomer B can be flowed through other channels, a monomer C can be flowed through still other channels, etc. In this manner, many or all of the reaction regions are reacted with a monomer before the channel block must be moved or the

5 substrate must be washed and/or reactivated. By making use of many or all of the available reaction regions simultaneously, the number of washing and activation steps can be minimized.

One of skill in the art will recognize that there are alternative methods of forming channels or otherwise protecting a portion of the surface of the substrate. For
10 example, according to some embodiments, a protective coating such as a hydrophilic or hydrophobic coating (depending upon the nature of the solvent) is utilized over portions of the substrate to be protected, sometimes in combination with materials that facilitate wetting by the reactant solution in other regions. In this manner, the flowing solutions are further prevented from passing outside of their designated flow paths.

15 The "spotting" methods of preparing nucleic acid libraries can be implemented in much the same manner as the flow channel methods. For example, a monomer A can be delivered to and coupled with a first group of reaction regions which have been appropriately activated. Thereafter, a monomer B can be delivered to and reacted with a second group of activated reaction regions. Unlike the flow channel
20 embodiments described above, reactants are delivered by directly depositing (rather than flowing) relatively small quantities of them in selected regions. In some steps, of course, the entire substrate surface can be sprayed or otherwise coated with a solution. In preferred embodiments, a dispenser moves from region to region, depositing only as much monomer as necessary at each stop. Typical dispensers include a micropipette to
25 deliver the monomer solution to the substrate and a robotic system to control the position of the micropipette with respect to the substrate, or an ink-jet printer. In other embodiments, the dispenser includes a series of tubes, a manifold, an array of pipettes, or the like so that various reagents can be delivered to the reaction regions simultaneously. Still other spotting methods are described in PCT/US99/00730.

30

Pin-Based Methods

Another method which is useful for the preparation of nucleic acid arrays and libraries involves "pin based synthesis." This method is described in detail in U.S. Pat. No. 5,288,514. The method utilizes a substrate having a plurality of pins or other

extensions. The pins are each inserted simultaneously into individual reagent containers in a tray. In a common embodiment, an array of 96 pins/containers is utilized.

Each tray is filled with a particular reagent for coupling in a particular chemical reaction on an individual pin. Accordingly, the trays will often contain different

5 reagents. Since the chemistry disclosed herein has been established such that a relatively similar set of reaction conditions may be utilized to perform each of the reactions, it becomes possible to conduct multiple chemical coupling steps simultaneously. In the first step of the process the invention provides for the use of substrate(s) on which the chemical coupling steps are conducted. The substrate is optionally provided with a spacer
10 having active sites. In the particular case of nucleic acids, for example, the spacer may be selected from a wide variety of molecules which can be used in organic environments associated with synthesis as well as aqueous environments associated with binding studies. Examples of suitable spacers are polyethyleneglycols, dicarboxylic acids, polyamines and alkenes, substituted with, for example, methoxy and ethoxy groups.
15 Additionally, the spacers will have an active site on the distal end. The active sites are optionally protected initially by protecting groups. Among a wide variety of protecting groups which are useful are Fmoc, Boc, t-butyl esters, t-butyl ethers, and the like. Various exemplary protecting groups are described in, for example, Atherton *et al.*, SOLID PHASE PEPTIDE SYNTHESIS, IRL Press (1989). In some embodiments, the spacer may
20 provide for a cleavable function by way of, for example, exposure to acid or base.

Libraries on Multiple Substrates

Bead Based Methods

Yet another method which is useful for synthesis of nucleic acid arrays
25 involves "bead based synthesis." A general approach for bead based synthesis is described in PCT/US93/04145 (filed Apr. 28, 1993).

For the synthesis of nucleic acids on beads, a large plurality of beads are suspended in a suitable carrier (such as water) in a container. The beads are provided with optional spacer molecules having an active site. The active site is protected by an
30 optional protecting group.

In a first step of the synthesis, the beads are divided for coupling into a plurality of containers. For the purposes of this brief description, the number of containers will be limited to three, and the monomers denoted as A, B, C, D, E, and F. The protecting groups are then removed and a first portion of the molecule to be

synthesized is added to each of the three containers (i. e., A is added to container 1, B is added to container 2 and C is added to container 3).

Thereafter, the various beads are appropriately washed of excess reagents, and remixed in one container. Again, it will be recognized that by virtue of the large

5 number of beads utilized at the outset, there will similarly be a large number of beads randomly dispersed in the container, each having a particular first portion of the monomer to be synthesized on a surface thereof.

Thereafter, the various beads are again divided for coupling in another group of three containers. The beads in the first container are deprotected and exposed to
10 a second monomer (D), while the beads in the second and third containers are coupled to molecule portions E and F respectively. Accordingly, molecules AD, BD, and CD will be present in the first container, while AE, BE, and CE will be present in the second container, and molecules AF, BF, and CF will be present in the third container. Each
15 bead, however, will have only a single type of molecule on its surface. Thus, all of the possible molecules formed from the first portions A, B, C, and the second portions D, E, and F have been formed.

The beads are then recombined into one container and additional steps such as are conducted to complete the synthesis of the polymer molecules. In a preferred embodiment, the beads are tagged with an identifying tag which is unique to the
20 particular nucleic acid or probe which is present on each bead. A complete description of identifier tags for use in synthetic libraries is provided in co-pending application Ser. No. 08/146,886 (filed Nov. 2, 1993).

Solid supports

25 Solid supports used in the present invention include any of a variety of fixed organizational support matrices. In some embodiments, the support is substantially planar. In some embodiments, the support may be physically separated into regions, for example, with trenches, grooves, wells and the like. Examples of supports include slides, beads and solid chips. Additionally, the solid supports may be, for example, biological,
30 nonbiological, organic, inorganic, or a combination thereof, and may be in forms including particles, strands, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates, and slides depending upon the intended use.

Supports having a surface to which arrays of nucleic acids are attached are also referred to herein as "biological chips". The support is preferably, silica or glass, and

can have the thickness of a microscope slide or glass cover slip. Supports that are transparent to light are useful when the assay involves optical detection, as described, e.g., in U.S. Patent No. 5,545,531. Other useful supports include Langmuir Blodgett film, germanium, (poly)tetrafluorethylene, polystyrene, (poly)vinylidenedifluoride,

5 polycarbonate, gallium arsenide, gallium phosphide, silicon oxide, silicon nitride, and combinations thereof. In one embodiment, the support is a flat glass or single crystal silica surface with relief features less than about 10 Angstroms.

The surfaces on the solid supports will usually, but not always, be composed of the same material as the substrate. Thus, the surface may comprise any
10 number of materials, including polymers, plastics, resins, polysaccharides, silica or silica based materials, carbon, metals, inorganic glasses, membranes, or any of the above-listed substrate materials. Preferably, the surface will contain reactive groups, such as carboxyl, amino, and hydroxyl. In one embodiment, the surface is optically transparent and will have surface Si-OH functionalities such as are found on silica surfaces. In other
15 embodiments, the surface will be coated with functionalized silicon compounds (see, for example, U.S. Patent No. 5,919,523).

Surface Density

The nucleic acid arrays described herein can have any number of nucleic
20 acid sequences selected for different applications. Typically, there may be, for example, about 100 or more, or in some embodiments, more than 10^5 or 10^8 . In one embodiment, the surface comprises at least 100 probe nucleic acids each preferably having a different sequence, each probe contained in an area of less than about 0.1 cm^2 , or, for example, between about $1/\text{mm}^2$ and $10,000/\text{mm}^2$, and each probe nucleic acid having a defined
25 sequence and location on the surface. In one embodiment, at least 1,000 different nucleic acids are provided on the surface, wherein each nucleic acid is contained within an area less than about 10^{-3} cm^2 , as described, for example, in U.S. Patent No. 5,510,270.

Arrays of nucleic acids for use in gene expression monitoring are described in PCT WO 97/10365, the disclosure of which is incorporated herein. In one
30 embodiment, arrays of nucleic acid probes are immobilized on a surface, wherein the array comprises more than 100 different nucleic acids and wherein each different nucleic acid is localized in a predetermined area of the surface, and the density of the different nucleic acids is greater than about 60 different nucleic acids per 1 cm^2 .

Arrays of nucleic acids immobilized on a surface which may be used also are described in detail in U.S. Patent No. 5,744,305, the disclosure of which is incorporated herein. As disclosed therein, on a substrate, nucleic acids with different sequences are immobilized each in a predefined area on a surface. For example, 10, 50, 60, 100, 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , or 10^8 different monomer sequences may be provided on the substrate. The nucleic acids of a particular sequence are provided within a predefined region of a substrate, having a surface area, for example, of about 1 cm^2 to 10^{-10} cm^2 . In some embodiments, the regions have areas of less than about 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6} , 10^{-7} , 10^{-8} , 10^{-9} , or 10^{-10} cm^2 . For example, in one embodiment, there is provided a planar, non-porous support having at least a first surface, and a plurality of different nucleic acids attached to the first surface at a density exceeding about 400 different nucleic acids/ cm^2 , wherein each of the different nucleic acids is attached to the surface of the solid support in a different predefined region, has a different determinable sequence, and is, for example, at least 4 nucleotides in length. The nucleic acids may be, for example, about 4 to 20 nucleotides in length. The number of different nucleic acids may be, for example, 1000 or more. In the embodiment where polynucleotides of a known chemical sequence are synthesized at known locations on a substrate, and binding of a complementary nucleotide is detected, and wherein a fluorescent label is detected, detection may be implemented by directing light to relatively small and precisely known locations on the substrate. For example, the substrate is placed in a microscope detection apparatus for identification of locations where binding takes place. The microscope detection apparatus includes a monochromatic or polychromatic light source for directing light at the substrate, means for detecting fluoresced light from the substrate, and means for determining a location of the fluoresced light. The means for detecting light fluoresced on the substrate may in some embodiments include a photon counter. The means for determining a location of the fluoresced light may include an x/y translation table for the substrate. Translation of the substrate and data collection are recorded and managed by an appropriately programmed digital computer, as described in U.S. Patent No. 5,510,270.

Non-Oxidizing Atmospheres

In the present invention, synthesis of the nucleic acid arrays using any of the techniques described herein is conducted in a non-oxidizing atmosphere. The use of such an atmosphere has been found to increase the yield of nucleic acid synthesis on a

solid support and to reduce any degradation of the resultant array. Accordingly, a non-oxidizing atmosphere is an atmosphere that will produce less than 30% degradation of a nucleic acid over a period of about 1 hour at ambient temperature (about 18 to about 25°C). More preferably, a non-oxidizing atmosphere produces less than 20%

5 degradation, still more preferably less than 10% degradation. Most preferably, a non-oxidizing atmosphere produces less than 5% degradation of a nucleic acid under the conditions noted above. The non-oxidizing atmosphere will preferably be an inert gas atmosphere (e.g., argon), a nitrogen atmosphere, or a filtered air atmosphere wherein the filter removes substantially all interfering amounts of ozone. Typically, a filtered air
10 atmosphere is filtered with a carbon filter and has a concentration of about 5 ppb or less ozone.

Depending on the scale of nucleic acid array preparation, the non-oxidizing atmosphere can be achieved by conducting the synthesis in a "glove box," a production facility or building equipped with air filters (e.g., carbon filters to remove
15 atmospheric ozone), or any facility (fume hood, reaction vessel, etc.) that can be equipped with a positive flow of a non-oxidizing atmosphere such as, for example, an inert gas, a nitrogen atmosphere, or filtered air.

Typically, the non-oxidizing atmosphere will have a concentration of ozone that is less than 10 ppb, more preferably less than 5 ppb and most preferably from
20 about 0 to 2 ppb.

Applications Using Nucleic Acid Arrays

The methods and compositions described herein may be used in a range of applications including biomedical and genetic research as well as clinical diagnostics.
25 Arrays of polymers such as nucleic acids may be screened for specific binding to a target, such as a complementary nucleotide, for example, in screening studies for determination of binding affinity and in diagnostic assays. In one embodiment, sequencing of polynucleotides can be conducted, as disclosed in U.S. Patent No. 5,547,839. The nucleic acid arrays may be used in many other applications including detection of genetic
30 diseases such as cystic fibrosis, diabetes, and acquired diseases such as cancer, as disclosed in U.S. Patent Application Ser. No. 08/143,312. Genetic mutations may be detected by sequencing by hybridization. In one embodiment, genetic markers may be sequenced and mapped using Type-IIs restriction endonucleases as disclosed in U.S. Patent No. 5,710,000.

Other applications include chip based genotyping, species identification and phenotypic characterization, as described in U.S. Patent Application Serial No. 08/797,812, filed February 7, 1997, and U.S. Application Serial No. 08/629,031, filed April 8, 1996. Still other applications are described in U.S. Patent No. 5,800,992.

5 Gene expression may be monitored by hybridization of large numbers of mRNAs in parallel using high density arrays of nucleic acids in cells, such as in microorganisms such as yeast, as described in Lockhart *et al.*, *Nature Biotechnology*, 14:1675-1680 (1996). Bacterial transcript imaging by hybridization of total RNA to nucleic acid arrays may be conducted as described in Saizieu *et al.*, *Nature*
10 *Biotechnology*, 16:45-48 (1998). Accessing genetic information using high density DNA arrays is further described in Chee, *Science* 274:610-614 (1996).

Still other methods for screening target molecules for specific binding to arrays of polymers, such as nucleic acids, immobilized on a solid substrate, are disclosed, for example, in U.S. Patent No. 5,510,270. The fabrication of arrays of polymers, such as
15 nucleic acids, on a solid substrate, and methods of use of the arrays in different assays, are also described in: U.S. Patent Nos. 5,677,195, 5,624,711, 5,599,695, 5,445,934, 5,451,683, 5,424,186, 5,412,087, 5,405,783, 5,384,261, 5,252,743 and 5,143,854; PCT WO 92/10092; and U.S. Application No. 08/388,321, filed February 14, 1995.

Devices for concurrently processing multiple biological chip assays are
20 useful for each of the applications described above (see, for example, U.S. Patent No. 5,545,531). Methods and systems for detecting a labeled marker on a sample on a solid support, wherein the labeled material emits radiation at a wavelength that is different from the excitation wavelength, which radiation is collected by collection optics and imaged onto a detector which generates an image of the sample, are disclosed in U.S.
25 Patent No. 5,578,832. These methods permit a highly sensitive and resolved image to be obtained at high speed. Methods and apparatus for detection of fluorescently labeled materials are further described in U.S. Patent Nos. 5,631,734 and 5,324,633.

In view of the technologies provided above, the present invention provides
30 in one preferred embodiment, a method of preparing a nucleic acid array on a support, wherein each nucleic acid occupies a separate predefined region of the support and the nucleic acids are synthesized using the steps:

(a) activating a region of the support;

(b) attaching a nucleotide to a first region, the nucleotide having a masked reactive site linked to a protecting group;

(c) repeating steps (a) and (b) on other regions of the support whereby each of the other regions has bound thereto another nucleotide comprising a masked

5 reactive site link to a protecting group, wherein the another nucleotide may be the same or different from that used in step (b);

(d) removing the protecting group from one of the nucleotides bound to one of the regions of the support to provide a region bearing a nucleotide having an unmasked reactive site;

10 (e) binding an additional nucleotide to the nucleotide with an unmasked reactive site;

(f) repeating steps (d) and (e) on regions of the support until a desired plurality of nucleic acids is synthesized, each nucleic acid occupying separate predefined regions of the support;

15 wherein each of steps (a) through (f) are carried out in a non-oxidizing atmosphere.

Preferably, the "activating" of step (a) is carried out using a channel block or photolithography technique, more preferably a photolithography technique. The "attaching" of step (b) is typically carried out using chemical means to provide a covalent
20 bond between the nucleotide and a surface functional group present in the first region. In some embodiments, the surface functional group will be a group present on a nucleotide or nucleic acid that is already attached to the solid support. For example, nucleic acid arrays can be prepared using a solid support having a surface coated with poly-A nucleic acids to provide suitable spacing between the surface of the support and the nucleic acids that will be used in subsequent hybridization assays. Accordingly, the "attaching" can be,
25 for example, by formation of a covalent bond between surface Si-OH groups and a group present on the first nucleotide of a nascent nucleic acid chain, or by formation of a covalent bond between groups present in a support-bound nucleic acid and a group present on the first nucleotide of a nascent nucleic acid. Typically, the groups present on
30 nucleic acids which are used in covalent bond formation are the 3'- or 5-hydroxyl groups in the sugar portion or the molecule, or phosphate groups attached thereto.

The nucleotides used in this and other aspects of the present invention will typically be the naturally-occurring nucleotides, derived from, for example, adenosine, guanosine, uridine, cytidine and thymidine. In certain embodiments, however, nucleotide

analogous or derivatives will be used (e.g., those nucleosides or nucleotides having protecting groups on either the base portion or sugar portion of the molecule, or having attached or incorporated labels, or isosteric replacements which result in monomers that behave in either a synthetic or physiological environment in a manner similar to the

5 parent monomer). The nucleotides will typically have a protecting group which is linked to, and masks, a reactive group on the nucleotide. A variety of protecting groups are useful in the invention and can be selected depending on the synthesis techniques employed. For example, channel block methods can use acid- or base-cleavable protecting groups to mask a hydroxyl group in a nucleotide. After the nucleotide is
10 attached to the support or growing nucleic acid, the protecting group can be removed by flowing an acid or base solution through an appropriate channel on the support.

Similarly, photolithography techniques can use photoremoveable protecting groups. Some classes of photoremoveable protecting groups include 6-nitroveratryl (NV), 6-nitropiperonyl (NP), methyl-6-nitroveratryl (MeNV), methyl-6-
15 nitropiperonyl (MeNP), and 1-pyrenylmethyl (PyR), which are used for protecting the carboxyl terminus of an amino acid or the hydroxyl group of a nucleotide, for example. 6-nitroveratryloxycarbonyl (NVOC), 6-nitropiperonyloxycarbonyl (NPOC), methyl-6-nitroveratryloxycarbonyl (MeNVOC), methyl-6-nitropiperonyloxycarbonyl (MeNPOC), 1-pyrenylmethyloxycarbonyl (PyROC), which are used to protect the amino terminus of
20 an amino acid are also preferred. Clearly, many photosensitive protecting groups are suitable for use in the present invention (see, U.S. Patent No. 5,489,678 and PCT WO 94/10128).

In addition, novel photoremoveable protecting groups such as 5'-O-pyrenylmethyloxy carbonyl (PYMOC) and methylnitropiperonyloxycarbonyl (MeNPOC)
25 have been described in the copending U.S. Patent Application Ser. No. 08/630,148, filed April 10, 1996, the contents of which are hereby incorporated by reference.

In addition to the above-described protecting groups, the present invention employs protecting groups, such as the 5'-X-2'-deoxythymidine 2-cyanoethyl 3'-N,N-diisopropylphosphoramidites in various solvents. In these protecting groups, X may
30 represent the following photolabile groups: ((α -methyl-2-nitropiperonyl)-oxy)carbonyl

(MeNPOC), ((Phenacyl)-oxy)carbonyl (PAOC), O-(9-phenylxanthen-9-yl) (PIXYL), and ((2-methylene-9,10-anthraquinone)-oxy)carbonyl (MAQOC).

Various methods for generating protected monomers have been described by the U.S. Patent No. 5,744,305, which is incorporated by reference. Detailed methods for using photoremovable protecting groups are described in the U.S. Patent No. 5,424,186, which is also hereby incorporated by reference.

The removal rate of the protecting groups depends on the wavelength and intensity of the incident radiation, as well as the physical and chemical properties of the protecting group itself. Preferred protecting groups are removed at a faster rate and with a lower intensity of radiation. For example, at a given set of conditions, MeNVOC and MeNPOC are photolytically removed faster than their unsubstituted parent compounds, NVOC and NPOC, respectively.

In addition to the above-described references, photocleavable protecting groups and methods of using such photocleavable protecting groups for polymer synthesis have been described in the copending applications 08/630,148 (filed April 10, 1996) and 08/812,005 (filed March 5, 1997) which are incorporated by reference herein.

Step (c) provides that steps (a) and (b) can be repeated to attach nucleotides to other regions of the solid support.

One of skill in the art will appreciate that steps (a) and (b) can be repeated a number of times to produce a solid support having a layer of attached nucleotides. Preferably, each attached nucleotide is in a preselected position.

In subsequent steps (d), (e) and (f), the protecting group is removed from one of the nucleotides to reveal a reactive site on the nucleotide. Thereafter, an additional nucleotide (optionally having a masked reactive site attached to a protecting group) is attached to the support-bound nucleotide. As above, these steps can be repeated to selectively attach an additional nucleotide to any of the support-bound nucleotides. Still further, the steps of deprotecting and attaching an additional nucleotide can be carried out on the newly added nucleotides to continue the synthesis of the nascent nucleic acid.

Each of the above steps is preferably carried out in a non-oxidizing atmosphere. Preferably, the non-oxidizing atmosphere is an inert gas, such as, for example, argon. In other preferred embodiments, the non-oxidizing atmosphere is filtered

air. In still other preferred embodiments, the non-oxidizing atmosphere is nitrogen or oxygen which has been purified to reduce the ozone concentration to 5 ppb or less.

~~In a further preferred embodiment, synthesis of the nucleic acids on the~~

5 support comprises the sequential steps of:

a) generating a pattern of light and dark areas by selectively irradiating at least a first area of a surface of a substrate, the surface comprising immobilized nucleotides on the surface, the nucleotides capped with a photoremovable protective group, without irradiating at least a second area of the surface, to remove the protective
10 group from the nucleotides in the first area;

b) simultaneously contacting the first area and the second area of the surface with a first nucleotide to couple the first nucleotide to the immobilized nucleotides in the first area, and not in the second area, the first nucleotide capped with the photoremovable protective group;

15 c) generating another pattern of light and dark areas by selectively irradiating with light at least a part of the first area of the surface and at least a part of the second area to remove the protective group in at least a part of the first area and at least a part of the second area;

d) simultaneously contacting the first area and the second area of the
20 surface with a second nucleotide to couple the second nucleotide to the immobilized nucleotides in at least a part of the first area and at least a part of the second area;

e) performing additional irradiating and nucleotide contacting and coupling steps so that a matrix array of at least 100 nucleic acids having different sequences is formed on the support;

25 with the proviso that steps (a) through (e) are performed in the non-oxidizing atmosphere, and the atmosphere preferably has an ozone concentration of from about 0 to about 5 ppb.

In this embodiment of the invention, the steps of generating patterns of light and dark areas, coupling nucleotides to specific areas, removing protecting groups
30 from the coupled nucleotides, and coupling additional nucleotides can all be carried out as described in, for example, U.S. Patent No. 5,510,270, with the added feature that these steps are performed in a non-oxidizing atmosphere, preferably having an ozone concentration of 5 ppb or less. More preferably, the non-oxidizing atmosphere is an inert gas (e.g., argon), filtered air, or a purified nitrogen or oxygen atmosphere.

EXAMPLES

In each of the examples below, the nucleic acid probe arrays were

5 prepared using photolithography and a silica wafer as the solid substrate. Preparation is typically on a 5 inch by 5 inch wafer which can be cut into 49 replicates of a probe array having about 400,000 distinct probe sequences, or 400 replicates of a probe array having about 50,000 distinct probe sequences. The density of the nucleic acid probes is about 1-10 picomoles per cm^2 .

10

Example 1

This example illustrates the post-synthesis degradation of various nucleic acid probes in unfiltered air.

15

In this example, a group of nucleic acid probes were synthesized on a silica wafer having a cleavable linker and subjected to ambient air, and 1000 ppb ozone. After 15 minutes, the probes were cleaved from the wafer and analyzed by HPLC. The probes included a mixed 16-mer, and homopolymers (8-mers) of poly-A, poly-T, poly-C and poly-G. Figure 1 provides the results after exposure to unfiltered air (Fig. 1A) and to

20

1000 ppb ozone (Fig. 1B).

As seen in Figure 1A, exposure of the homopolymers to ambient air having an ozone concentration of about 5-25 ppb results in significant degradation of the poly-T probe, with only minor amounts of degradation for the poly-C, poly-A and poly-G probes.

25

In Figure 1B, homopolymer probes having the same sequences as those in Figure 1A, and a mixed 16-mer were exposed to 1000 ppb ozone for 15 min. After the exposure, the mixed nucleic acid was degraded to about 20% of its initial amount, while poly-A and poly-G were essentially unchanged. Poly-T was completely degraded at 15 min, while poly-C was also substantially degraded (about 35% remained).

30

Example 2

This example illustrates the decay of a support-bound nucleic acid as a function of time when exposed to an atmosphere of 20 ppb ozone, and as a function of total ozone concentration.

An array of identical 16-mers having the sequence 3'-CGACATTTACAGTAAG-5' was prepared on a silica wafer with a cleavable linking group. The array was exposed to an atmosphere of clean, dry air (Air Products, Inc.) with a controlled ozone concentration (20 ppb, using an ozone generator, Environics Series 100)), for up to 24 hours. At time points of 0, 8, 16 and 24 hours, portions of the array were cleaved and subjected to HPLC analysis to determine the amount of nucleic acid degradation. The results are shown in Figure 2. In this figure, the 16-mer is about 50% degraded ($t_{1/2}$) at 7 hours exposure. Additionally, the observed rate is similar to the rate of decay in ambient air having 30 ppb ozone concentration (results not shown).

Similarly, Figure 3 shows the nucleic acid decay as a function of total ozone exposure (concentration x time; and normalized to an initial reading of zero exposure). The solid diamonds in Figure 3 are derived from a controlled ozone/CDA standard curve while the open square represents a sample exposed to ambient air, which also falls on the standard curve. This data implicates ozone as the agent responsible for nucleic acid degradation from unfiltered air.

Example 3

This example illustrates the protection of nucleic acid arrays that can be achieved when the arrays are stored under a non-oxidizing atmosphere.

Presynthesized nucleic acids (16-mers as described above) on silica wafers were exposed to a variety of environmental conditions for periods of from about 12 to about 15 hours. The results are normalized to a control nucleic acid having no exposure to the environment.

As shown in Figure 4, the nucleic acid exposed to ultrapure air exhibited little detectable degradation after 15 hr. Similar results were obtained for the exposure of a probe to ambient air that had been filtered through activated carbon to remove ozone (Balston filter, Whatman Scientific). Exposure of the array to pure nitrogen resulted in a reduction of about 15% in surface density, while exposure to pure oxygen resulted in a

reduction of about 20% in surface density. When the nucleic acid was subjected to ambient air for a period of 12 hr (ozone concentration of about 10-30 ppb), a reduction of 50% of the surface density resulted. Substantial degradation (almost 85%) was seen when the nucleic acid was exposed to ultrapure air having an added concentration of 20 (+/- 15) ppb ozone (Enviro-nics Series 100 ozone generator).

Example 4

This example illustrates the yield increase that can be achieved in a backside photolithographic technique that reduces exposure of the nascent nucleic acid to ambient air.

In this example, a 16-mer poly-A nucleic acid was prepared by two techniques. In the first approach, the 16-mer was prepared using photolithography and a monomer by monomer construction involving both deprotection and coupling steps (MeNPOC-protected adenine monomers). Irradiation of the substrate was carried out using a photolithographic mask and frontside illumination. In this method, the developing probe is left exposed to ambient air between the base addition and photolysis steps (to remove the MeNPOC groups). In the second approach, the 16-mer was also prepared using photolithography, but illumination was from the backside of the silica support. When the wafer is illuminated from the backside, it does not have to be removed from the protected environment of the flowcell (containing the monomers) at any time during synthesis. Accordingly, there is no exposure to an ambient atmosphere during a backside synthesis procedure.

Figure 5 is a bar graph which indicates the increase in synthesis yield that is achieved using backside illumination. As seen in the figure, frontside synthesis of a poly-A 16-mer results in a synthesis yield of about 6%, while backside synthesis provides a yield of almost 15%.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.